unibz
Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

Too much or too little?
New tools for the static CCE Estimator.

Jan Ditzen

Free University of Bozen-Bolzano, Bozen, Italy

www.jan.ditzen.net, jan.ditzen@unibz.it

https://janditzen.github.io/xtdcce2/

May 9, 2024

Motivation I

- In panel time series models with $(N, T) \Rightarrow \infty$ cross-section dependence via common factors (or unobserved heterogeneity) is likely to occur.
- Often modelled using interactive fixed effects

$$y_{i,t} = \beta_i x_{i,t} + u_{i,t} \quad (1)$$

$$u_{i,t} = \gamma_{1,i} f_{1,t} + ... + \gamma_{M,i} f_{M,t} + \vartheta_{i,t} \quad (2)$$

- where we have M interactions between unit specific effects and time effects.
- $\gamma_{m,i}$ is the factor loading of factor $f_{m,t}$.

Motivation II

- The explanatory variables can also consist of a factor structure:

$$x_{i,t} = \gamma_{x,1,i} f_{x,1,t} + ... + \gamma_{M_x,i} f_{M_x,t} + \epsilon_{i,t}$$

- Assumption: $corr(\vartheta_{i,t}, \epsilon_{i,t}) = 0$!
- This setting poses two challenges:
  1. Correlation across units, cross-sectional dependence
  2. If the factors in $x_{i,t}$ and $u_{i,t}$ overlap, the observables and unobservables are correlated
- Popular estimator: Common Correlated Effects Estimator (Pesaran, 2006)

(Static) Common Correlated Effects

- Pesaran (2006) proposes to span the space of the common factors using cross-section averages (CSA) in a static model:

$$y_{i,t} = \beta_i x_{i,t} + \psi_{x,i}\bar{x}_t + \psi_{y,i}\bar{y}_t + \epsilon_{i,t}$$

  where $\bar{x}_t = 1/N \sum_{i=1}^{N} x_{i,t}$ and $\bar{y}_t = 1/N \sum_{i=1}^{N} y_{i,t}$ are the cross-section averages.
- Can be combined with a pooled or mean group estimator.
- Estimator is $\sqrt{N}$ consistent.
- Proved to be versatile in many conditions (Kapetanios et al., 2011; Chudik et al., 2011; Westerlund, 2018).
- Extended to dynamic models (Chudik and Pesaran, 2015) and estimation of long run coefficients (Chudik et al., 2016).

Static CCE

Stata Implementations

- First implementation in `xtmg` (Eberhardt, 2012).
- Also implemented in `xtdcce2` (Ditzen, 2018, 2021):
  - Static and dynamic models
  - Estimation of long run coefficients
  - Bootstrapping
  - Different types of cross-section averages
  - Estimation of degree and testing for cross-section dependence
  - Various `estat` and `predict` functions
- Syntax:

`xtdcce2` *depvar indepvars* [*if*] , ... *cr(varlist , options)*

  - `cr()` defines variables added as cross-section averages.

Static CCE

New developments

- In “early” years discussion on CCE focused on validity under stationary factors, autocorrelated factors, strong and semi strong cross-section dependence and dynamic models.
- A recurring topic is also bootstrapping.
- In past years discussion on what “spanning factor space” actually means intensified:
    1. Regularized CCE
    2. When does the “Rank Condition” hold?
    3. Information Criteria to select CSA.

Regularized CCE

- Large number of cross-section averages might only contain limited information, inducing non-trivial bias for pooled and mean group CCE, see Karabiyik et al. (2017).
- Juodis (2022) suggest rCCE approach:
  - Calculate cross-section averages (CSA)
  - Estimate number of common factors in CSA, $\hat{m}$, using ER or GR from Ahn and Horenstein (2013)
  - Replace cross-section averages with the first $\hat{m}$ eigenvectors of CSA.
- Requires bootstrapping.
- Disadvantage: approach sensitive to estimation of factors and only for static panels.

Regularized CCE

`xtdcce2, ... cr(rcce)`

`xtdcce2` *depvar indepvars* [*if*] , ... *cr(varlist* , *rcce[(options)])*

- `options` are:
  - `criterion(er|gr)` specifies criterion to estimate number of common factors using the ER or GR criterion from Ahn and Horenstein (2013)
  - `scale` scales cross-section averages
  - `npc(real)` specifies number of eigenvectors without estimating it.
- Bootstrap is not automatically performed.
- Number of factors estimated based on `xtnumfac` (Ditzen and Reese, 2023).
- Unbalanced panels supported, then missing values in CSA are imputed.

Regularized CCE

Example

- Dell et al. (2012) investigate effect of temperature ($wtem_{i,t}$) and precipitation ($wpre_{i,t}$) on economic growth ($g_{i,t}$).
- Balanced panel of 89 countries and over 42 years (1962-2003):[1]

$$g_{i,t} = \mu_i + \beta_{1,i} wtem_{i,t} + \beta_{2,i} wpre_{i,t} + u_{i,t}$$

- Estimated number of common factors (output shortened):

```
. xtnumfac g wtem wpre , stand(5)
 N    =      3738                    T      =        42
 N_g  =        89                    vars.  =         3
-------------------------------------------------------
 IC         |  # factors   |  IC       |  # factors
-------------------------------------------------------
...
 ER         |      1       |  GR       |      1
...
-------------------------------------------------------
8 factors maximally considered.
...
ER, GR from Ahn and Horenstein (2013)
...
```

[1] This is only an example!

Regularized CCE

Example MG (no CSA)

```
xtdcce2 g wtem wpre, cr(g wtem wpre)
(Dynamic) Common Correlated Effects Estimator - Mean Group

Panel Variable (i): cc_num                  Number of obs      =      3738
Time Variable (t): year                     Number of groups   =        89

Degrees of freedom per group:               Obs per group (T) =        42
 without cross-sectional averages   = 39
 with cross-sectional averages      = 36
Number of                                   F(534, 3204)       =      0.40
 cross-sectional lags               0 to 0  Prob > F           =      1.00
 variables in mean group regression = 178   R-squared          =      0.94
 variables partialled out           = 356   R-squared (MG)     =      0.10
                                            Root MSE           =      5.04
                                            CD Statistic       =      0.79
                                               p-value         =    0.4322
-------------------------------------------------------------------------------
           g|     Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
------------+------------------------------------------------------------------
 Mean Group:|
        wtem|  -.492719   .3991216    -1.23    0.217     -1.274983    .289545
        wpre|  .0794217   .0863426     0.92    0.358     -.0898066   .2486501
-------------------------------------------------------------------------------
Mean Group Variables: wtem wpre
Cross Sectional Averaged Variables: g wtem wpre
Heterogenous constant partialled out.
```

- `cr(g wtem wprec)` adds cross-section averages of g, wtem and wprec.
- No strong cross-section dependence left in residuals.
- No coefficients significant.
- Space of one common factor spanned by 3 cross-section averages.

Regularized CCE

Example

```
. xtdcce2 g wtem wpre, cr(g wtem wpre, rcce(npc(1)))
(Dynamic) Common Correlated Effects Estimator - Mean Group
Panel Variable (i): cc_num                          Number of obs      =      3738
Time Variable (t): year                             Number of groups   =        89
Degrees of freedom per group:                       Obs per group (T)  =        42
 without cross-sectional averages   = 39
 with cross-sectional averages      = 36
Number of                                           F(534, 3204)       =      0.44
 cross-sectional lags               = 0             Prob > F           =      1.00
 variables in mean group regression = 178           R-squared          =      0.93
 variables partialled out           = 356           R-squared (MG)     =     -0.04
                                                    Root MSE           =      5.43
                                                    CD Statistic       =     18.31
                                                       p-value         =    0.0000
-------------------------------------------------------------------------------
          g |     Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
------------+------------------------------------------------------------------
 Mean Group:|
       wtem | -.8734534   .2471956    -3.53    0.000    -1.357948   -.3889591
       wpre |  .1390197   .0975041     1.43    0.154    -.0520847    .3301241
-------------------------------------------------------------------------------
Mean Group Variables: wtem wpre
1 Regularized Cross-Section Averages from variables:
 g wtem wpre
Heterogenous constant partialled out.
```

- `cr(... , rcce(npc(1))` employs rCCE estimator with 1st eigenvalue instead of CSA.
- Some strong cross-section dependence left in residuals.
- Temperature significant.
- `xtdcce2` can also estimate number of common factors in CSA Example.

Static CCE

The Rank Condition

- Key for consistent estimation is the rank condition.
- Rank condition implies that - loosely speaking(!!) - rank of average factor loadings has to be larger or equal than rank of factors.
- Karabiyik et al. (2017, 2021) show that if rank condition fails, CCE is inconsistent!
- Implies that adding CSAs with zero loadings, might lead to problems!
- In empirical practice: more CSA required than factors.

Rank Condition Classifier

- Reminder: Rank of unobserved factors (f), m, is not larger than the rank of unobserved average factor loadings (γ), g.
- Problem: both are unobserved! DeVos et al. (2024) propose a indicator if the rank conditions holds:

$$\widehat{RC} = 1 - I(\hat{g} < \hat{m}) \tag{3}$$

  where $\hat{m}$ is the rank of the cross/product of the observed data estimated by ER or GR from Ahn and Horenstein (2013) and $\hat{g}$ is rank of the unobserved factor loadings estimated from cross-section averages.
- If $\widehat{RC} = 1$, rank condition holds.
- Requires bootstrap to estimate variance of rank estimator of factor loadings.
- Consistency depends on fixed T. Trick: bound dimension with shrinkage.
- Indicator only valid for static panels.

Rank Condition Classifier

`xtdcce2, ... cr(rccl)`

`xtdcce2` *depvar indepvars* [*if*] , ... *cr(varlist , rcclassifier[(options)])*

- `options` are:
  - `er|gr` specifies criterion to estimate number of common factors using the ER or GR criterion from Ahn and Horenstein (2013)
  - `standardize(integer)` standardize data prior to estimation of number of common factors.
  - `replications(integer)` sets number of replication for bootstrapping variance of the rank estimator of the unobserved matrix of average factor loadings.
  - `randomshrinkage` Instead of fold-over matrix, use matrix with entries drawn from random normal distribution.
  - `noshrinkage` No shrinkage.
- Number of factors estimated based on `xtnumfac` (Ditzen and Reese, 2023).

Rank Condition Classifier

Example[2]

```
. xtdcce2 g wtem wpre, cr(g wtem wpre, rccl)
(Dynamic) Common Correlated Effects Estimator - Mean Group

Panel Variable (i): cc_num                          Number of obs      =      3738
Time Variable (t): year                             Number of groups   =        89

Degrees of freedom per group:                       Obs per group (T) =        42
 without cross-sectional averages    = 39
 with cross-sectional averages       = 36
Number of                                           F(534, 3204)       =      0.40
 cross-sectional lags                0 to 0         Prob > F           =      1.00
 variables in mean group regression = 178           R-squared          =      0.94
 variables partialled out           = 356           R-squared (MG)     =      0.10
                                                    Root MSE           =      5.04
                                                    CD Statistic       =      0.79
                                                       p-value         =    0.4322
-------------------------------------------------------------------------------
          g |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+------------------------------------------------------------------
 Mean Group:|
       wtem |   -.492719   .3991216   -1.23   0.217    -1.274983     .289545
       wpre |   .0794217   .0863426    0.92   0.358    -.0898066    .2486501
-------------------------------------------------------------------------------
Mean Group Variables: wtem wpre
Cross Sectional Averaged Variables: g wtem wpre
Heterogenous constant partialled out.
-------------------------------------------------------------
Test for Rank Condition (De Vos, Everaert and Sarafidis, 2024)
-------------------------------------------------------------
       RC      |      Estimated
  (1-I(p<m))*  |   Rank    # Factors
---------------+---------------------------------------------
      1        |    2         1
-------------------------------------------------------------
* RC=1 indicates rank condition holds.
```

- `cr(... , rccl)` requests calculation or rank condition classifier.
- More common factors in CSA than in unobserved factors.

[2]Unpublished beta version of `xtdcce2` version 4.7. Results may change.

Information Criteria and CCE

- Long time no theory for use of information criteria and CCE
- Information criteria can be used for 2 purposes:
  - Selection of cross-section averages
  - Lag selection in dynamic models
- 1) received recently attention:
  - DeVos et al. (2024) propose a sequential method to identify the set of relevant cross-section averages.
  - Margaritella and Westerlund (2023) propose a similar criteria as in Bai and Ng (2002); Bai (2009).
- **No** guidance on **lag selection** in dynamic models - left for research!

Information Criteria and CCE

- Margaritella and Westerlund (2023) propose 4 criteria to identify the optimal set of cross-section averages:

$$IC_1(M) = ln\hat{\sigma}^2\left(\hat{F}_M\right) + m\frac{N+T}{NT}ln\left(\frac{NT}{N+T}\right), \qquad IC_2(M) = ln\hat{\sigma}^2\left(\hat{F}_M\right) + m\frac{N+T}{NT}ln\left(C_{NT}^2\right)$$

$$PC_1(M) = \hat{\sigma}^2\left(\hat{F}_M\right) + m\hat{\sigma}^2\left(\hat{F}_{\bar{M}}\right)\frac{N+T}{NT}ln\left(\frac{NT}{N+T}\right), \quad PC_2(M) = \hat{\sigma}^2\left(\hat{F}_M\right) + m\hat{\sigma}^2\left(\hat{F}_{\bar{M}}\right)\frac{N+T}{NT}ln\left(C_{NT}^2\right)$$

  with $\hat{\sigma}^2(\hat{F}_M)$ the error variance from a CCE estimation with m CSA and $\left(\hat{F}_{\bar{M}}\right)$ is the error variance with the full set of CSA.
- Optimal number of CSA is then $\hat{M} = argminIC(M)$.
- Integrated in `xtdcce2` as `estat ic`.
- Caveat: can only be applied to static models.

Information Criteria and CCE

Example[3]

```
. estat ic
IC from Margaritella & Westerlund (2023)
```

Model	IC1	IC2	PC1	PC2
1	6.79*	6.59*	.	.

- ICs alone are not informative.
- Option `sequential` compares all permutations of CSA.

[3] Unpublished beta version of `xtdcce2` version 4.7. Results may change.

Information Criteria and CCE

Example[4]

```
. estat ic, seq
Running 7 combinations of cross-section averages:
. . . . . . .
IC from Margaritella & Westerlund (2023)
```

Model	IC1	IC2	PC1	PC2
1	6.95	6.88	1040	1040
2	6.95	6.88	1043	1043
3	6.86	6.79	948	948
4	6.91	6.77	1001	1001
5	6.82	6.68	909	909
6	6.82	6.69	918	918
7	6.79*	6.59*	884*	884*

```
* indicates minimum.
Cross Section Averages:
 Model 1: wpre
 Model 2: wtem
 Model 3: g
 Model 4: wtem wpre
 Model 5: g wpre
 Model 6: g wtem
 Model 7: g wtem wpre (Main Model)
Click on Model to run in xtdcce2.
```

- Model with 3 CSA has lowest IC.
- Possible to run models directly in Stata.
- Many combinations possible which can take time.

[4] Unpublished beta version of `xtdcce2` version 4.7. Results may change.

Summary

- New developments in the CCE literature on **static** models:
  - Regularized CCE (rCCE)
  - Rank condition classifier
  - Information criteria to select CSA
- rCCE available with `xtdcce2` version 4.0.
- Rank condition classifier and IC will be available with version 4.7.
- How to install?

```
net install xtdcce2 , from("https://janditzen.github.io/xtdcce2/")
```

- More info:

jan.ditzen.net

GitHub

References I

Ahn, S. C., and A. R. Horenstein. 2013. Eigenvalue Ratio Test for the Number of Factors. Econometrica 81(3): 1203–1227.

Bai, J. 2009. Panel Data Models With Interactive Fixed Effects. Econometrica 77(4): 1229–1279.

Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. Econometrica 70(1): 191–221.

Chudik, A., K. Mohaddes, M. H. Pesaran, and M. Raissi. 2016. Long-Run Effects in Large Heterogeneous Panel Data Models with Cross-Sectionally Correlated Errors. In Essays in Honor of Aman Ullah (Advances in Econometrics, Volumn 36), ed. R. C. Hill, G. GonzÁlez-Rivera, and T.-H. Lee, 85–135. Emerald Group Publishing Limited.

References II

Chudik, A., and M. H. Pesaran. 2015. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188(2): 393–420.

———. 2019. Mean group estimation in presence of weakly cross-correlated estimators. Economics Letters 175: 101–105.

Chudik, A., M. H. Pesaran, and E. Tosetti. 2011. Weak and strong cross-section dependence and estimation of large panels. The Econometrics Journal 14(1): C45–C90.

Dell, B. M., B. F. Jones, and B. A. Olken. 2012. Temperature Shocks and Economic Growth: Evidence from the Last Half Century. American Economic Journal: Macroeconomics 4(3): 66–95.

References III

DeVos, I., G. Everaert, and V. Sarafidis. 2024. A method to evaluate the rank condition for CCE estimators A method to evaluate the rank condition for CCE estimators. Econometric Reviews 0(0): 1–34. URL https://doi.org/10.1080/07474938.2023.2292383.

Ditzen, J. 2018. Estimating Dynamic Common-Correlated Effects in Stata. The Stata Journal 18(3): 585–617.

———. 2021. Estimating long run effects and the exponent of cross-sectional dependence: an update to xtdcce2. The Stata Journal 21(3): 687–707. URL https://ideas.repec.org/p/bzn/wpaper/bemps81.html.

Ditzen, J., and S. Reese. 2023. A battery of estimators for the number of common factors in time series and panel data models. The Stata Journal 23(2): 438–454. URL https://journals.sagepub.com/doi/abs/10.1177/1536867X231175305.

References IV

Eberhardt, M. 2012. Estimating panel time series models with heterogeneous slopes. The Stata Journal 12(1): 61–71.

Juodis, A. 2022. A regularization approach to common correlated effects estimation. Journal of Applied Econometrics (October 2021): 788–810.

Kapetanios, G., M. H. Pesaran, and T. Yamagata. 2011. Panels with non-stationary multifactor error structures. Journal of Econometrics 160(2): 326–348.

Karabiyik, H., S. Reese, and J. Westerlund. 2017. On the role of the rank condition in CCE estimation of factor-augmented panel regressions. Journal of Econometrics 197(1): 60–64. URL http://dx.doi.org/10.1016/j.jeconom.2016.10.006.

Karabiyik, H., J. Westerlund, and A. Juodis. 2021. On the Robustness of the Pooled CCE Estimator. Journal of Econometrics 220(2): 325–348.

References V

Margaritella, L., and J. Westerlund. 2023. Using Information Criteria to Select Averages in CCE. The Econometrics Journal .

Pesaran, M. H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4): 967–1012.

Pesaran, M. H., and R. Smith. 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68(1): 79–113.

Westerlund, J. 2018. CCE in panels with general unknown factors. Econometrics Journal 21(3): 264–276.

MG with no CSA back

```
. xtdcce2 g wtem wpre, nocross
(Dynamic) Common Correlated Effects Estimator - Mean Group

Panel Variable (i): cc_num                         Number of obs      =      3738
Time Variable (t): year                            Number of groups   =        89

Degrees of freedom per group:                      Obs per group (T)  =        42
 without cross-sectional averages    = 39
 with cross-sectional averages       = 39
Number of                                          F(267, 3471)       =      0.87
 cross-sectional lags                none          Prob > F           =      0.93
 variables in mean group regression = 178          R-squared          =      0.94
 variables partialled out            = 89          R-squared (MG)     =      0.01
                                                   Root MSE           =      5.28
                                                   CD Statistic       =     22.42
                                                      p-value         =    0.0000
------------------------------------------------------------------------------
           g |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Mean Group: |
        wtem |  -.6109652   .2163647   -2.82   0.005     -1.035032   -.1868981
        wpre |   .1805907   .0884139    2.04   0.041      .0073027    .3538786
------------------------------------------------------------------------------
Mean Group Variables: wtem wpre
Heterogenous constant partialled out.
```

rCCE - number of factors estimated back

```
. xtdcce2 g wtem wpre, cr(g wtem wpre, rcce)
(Dynamic) Common Correlated Effects Estimator - Mean Group

Panel Variable (i): cc_num                      Number of obs      =      3738
Time Variable (t): year                         Number of groups   =        89

Degrees of freedom per group:                   Obs per group (T)  =        42
 without cross-sectional averages     = 39
 with cross-sectional averages        = 36
Number of                                       F(534, 3204)       =      0.47
 cross-sectional lags                 = 0       Prob > F           =      1.00
 variables in mean group regression = 178       R-squared          =      0.93
 variables partialled out             = 356     R-squared (MG)     =      0.07
                                                Root MSE           =      5.11
                                                CD Statistic       =      0.43
                                                   p-value         =    0.6704
-------------------------------------------------------------------------------
           g|     Coef.    Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------------------------------------------------------------------------
 Mean Group:|
        wtem| -.0618736   .3554097    -0.17    0.862     -.7584638   .6347165
        wpre|  .1070665    .091807     1.17    0.244     -.0728719   .2870048
-------------------------------------------------------------------------------
Mean Group Variables: wtem wpre
2 Regularized Cross-Section Averages from variables:
 g wtem wpre
Heterogenous constant partialled out.
```

Information Criteria and CCE

Example

```
. estat ic, model((g wtem wpre) (g) (g wtem))
Running 3 combinations of cross-section averages:
. . .
IC from Margaritella & Westerlund (2023)
```

Model	IC1	IC2	PC1	PC2
1	6.79*	6.59*	885*	885*
2	6.86	6.79	950	950
3	6.82	6.69	918	918

```
* indicates minimum.
Cross Section Averages:
 Model 1: g wtem wpre (Main Model)
 Model 2: g
 Model 3: g wtem
Click on Model to run in xtdcce2.
```

- Compare specific models.
- Syntax: `model( (model1) (model2) ... ).`
- `(model1)` is the reference model.

Static CCE

Mean Group Estimator

- Main contributions: Pesaran and Smith (1995); Pesaran (2006); Chudik and Pesaran (2019)
- $\hat{\beta}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i$
- Asymptotic variance estimator $V(\hat{\beta}_{MG}) = \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\hat{\beta}_i - \hat{\beta}_{MG}\right)^2$
- Individual coefficients asymptotically normal with $(N, T) \xrightarrow{j} \infty$ with no particular order. Rank condition requires $\sqrt{T}/N \rightarrow 0$.
- Mean group asymptotically unbiased if $N \rightarrow \infty$ and T fixed and $(N, T) \xrightarrow{j} \infty$. Also needs $T > K$.

Static CCE

Pooled Estimator

- Main contribution: Pesaran (2006)
- Estimate β_p directly with the condition $\beta_i = \beta_p$.
- Various variance estimators, such as $V(\hat{\beta}_p)_{np} = f(\hat{\beta}_i, \hat{\beta}_{MG}, \tilde{X}'\tilde{X})$ or $V(\hat{\beta}_p)_{hac} = f(\hat{\beta}_p, \tilde{X}'\tilde{X}, \hat{\epsilon}_{i,t})$.
- Depending on the estimator, we need to make sure that the residuals are cross-section dependence free!
- Asymptotically normal with $(N, T) \xrightarrow{j} \infty$ with no particular order.